S-E-C-R-E-T

UNITED STATES INTELLIGENCE BOARD

COMMITTEE ON DOCUMENTATION

5 August 1963

MEMORANDUM FOR: Inspector General of CIA

SUBJECT: Coordinated CODIB Contribution on Objective 3

REFERENCE: Memorandum from DDCI, Subject: United States
Foreign Intelligence Objectives, Top Secret,
dated 8 July 1963

There are forwarded herewith 15 copies of the coordinated CODIB

response on Objective 3 as requested in the referenced memorandum.

25X1

John K. Vance
Acting Chairman

S-E-C-R-E-T

GROUP I
Excluded from automatic
downgrading and
declassification

S-E-C-R-E-T

5 August 1963

PROGRAMS AND PLANS FOR ACTION TO ACCOMPLISH
UNITED STATES FOREIGN INTELLIGENCE OBJECTIVE NUMBER 3

INFORMATION PROCESSING TECHNIQUES

I. OBJECTIVE NO. 3

That research be intensified to determine the usefulness of data processing techniques, including mechanized title or summary sentence permutation, to facilitate review and assessment of the great volume of material that must be dealt with in the intelligence community.

II. THE PROBLEM IN PERSPECTIVE

As of today, only man possesses the intelligence and insight to ask questions, to exercise judgment, and to recognize a good solution when it turns up. Modern information processing equipment and techniques, however, can be of substantial assistance to an intelligence analyst by drastically reducing the amount of time and effort required to search for the information he desires, to collate it with other pertinent information, to update it, to rearrange it in different ways to see if any patterns are developing, and to perform various types of mathematical computations and logical structuring. The machine will never replace the intelligence analyst, but, if skillfully used, it can provide an extremely powerful tool of analysis.

Research and development work in intelligence automation is currently at a high level and is expected to increase in the years immediately ahead. The rate of progress,

GROUP I
S=E-C-R-E-T   Excluded from automatic
downgrading and
declassification

however, will be governed more by availability of first-rate research workers than by the supply of funds.

One of the fundamental difficulties in applying automation to intelligence problems is that computer equipment in use today is basically designed for the processing of numerical data. Yet, most intelligence ADP efforts are primarily concerned with the receipt, processing, storage, and retrieval of non-numerical data.

Non-numerical information processing deals with the manipulation of symbols, meanings and decisions in an essentially qualitative and descriptive manner. As opposed to the processing of numerical data which is characterized by its fundamental reliance on arithmetic, the processing of non-numerical data includes structuring, ordering, indexing, searching, sorting or rearranging, merging, updating and the like. Present techniques for performing these tasks are generally not formalized and not reducible to a consistent, logical notation. In short, the task of defining the problem so that it can be programmed on the computer is much more difficult in non-numerical than in numerical processing.

III. OPERATIONAL AND DEVELOPMENT PROGRAMS

The intelligence community is presently organized so as to be responsive to changes as a result of the current technological revolution in information handling. The mechanism within the Intelligence Community for effecting improvement in intelligence handling techniques is exemplified by the DOD. Here each of the three military departments and the NSA have research and development organizations. DIA, in turn, develops intelligence research and development requirements for the DOD

Both NSA and DIA are responsible for ensuring that the projects of the military departments are operationally sound and meet operational requirements in the cryptologic and non-cryptologic intelligence communities, respectively. The Director of Defense Research and Engineering then coordinates and guides all of these efforts for the Secretary of Defense. Finally, DDR&E, NSA, DIA, State Dept., FBI, the Military Departments and CIA have representatives on the Committee on Documentation (CODIB) of the United States Intelligence Board. CODIB, which is chaired by CIA, provides central coordination and leadership for intelligence data handling matters. Problems are brought up by members for review and guidance and information is disseminated about on-going activities. For example, having determined the need for a survey of community information processing capabilities, CODIB organized the Staff for Community Information Processing Study (SCIPS) to perform the survey with group members supplied by CODIB representative agencies under CIA direction. Also, CODIB identified the problem of compromising emanations from operating flexowriter-type equipment and is resolving the problem on a community-wide basis through development of standard operating specifications, design and production of equipment and operating usage agreements.

During the past 17 years, as the Cold War has spread to additional areas and as communist security measures have tightened, the U. S. Intelligence Community has been forced to rely increasingly on mass collection techniques. This has resulted in enormous volumes of electronic, photographic, acoustic, and hard copy information which have inundated our manual processing resources.

Manual methods of processing have proved to be incapable of exploiting thoroughly the volume and type of raw data collected in the following fields: ELINT; targeting; photography; current missile, air, naval and ground forces activity; merchant shipping activity; and biographic material. Because of this, in the early 1950's efforts began in the Intelligence Community to apply punch card techniques to certain intelligence files, primarily in the targeting and document storage and retrieval areas. Each year since then, APD has been applied to a growing number of intelligence tasks. At present, operational and development programs are underway in the following areas:

A. Document Storage and Retrieval

In 1958 the first automated document storage and retrieval system (MINICARD) began to be implemented in Air Force Intelligence. Today a number of intelligence organizations possess machine supported document storage and retrieval systems tailored to their specific requirements.

CIA operates a central information control system in which about 275,000 incoming items per year are indexed according to a comprehensive subject classification scheme developed by the Intelligence Community. The index record for each document is stored on punch cards and the output to the intelligence analyst in response to a search query is a bibliography of document titles prepared by partially automated means. The information reports themselves are recorded on microfilm and mounted in aperture cards from which reproductions are readily obtained in response to requests.

The Air Force has conducted experiments at SAC, NORAD and the Foreign Technology Division (FTD) in a very promising indexing system for intelligence

documents called "Key Word in Context (KWIC) Indexing." For many years, efforts have been underway to develop automatic indexing techniques based on key words. The problem has been to combine key words into meaningful key terms, since an effective indexing system is most frequently based on terms rather than on single words. One of the most promising approaches now appears to be the KWIC system by means of which it is possible to retrieve a listing of all appearances of a given word in intelligence documents, together with the words appearing immediately before and after that word in each appearance. Thus, by seeing the word in context one can more readily find the document he is searching for. Such a system will soon become operational on a limited scale in the DIA Current Intelligence and Indications Center.

NSA has been conducting experiments in capturing titles of reports automatically and preparing therefrom permuted word indexes. The first results of this effort have become available recently and appear to be very encouraging. (NSA's applications of EDP techniques are otherwise in areas not pertinent to this paper.)

The Department of State is currently developing a pilot system in the area of Cuban affairs based on a computer and designed to produce permuted subject indexes for the analyst. At present, sentences describing the contents of approximately 5,000 documents of an 88,000 document library on Cuba have been prepared and machine input is expected to begin shortly.

In CIA a major systems study was initiated in 1962 to investigate the application of computers to the Agency's central document storage and retrieval system. The study is now completing an investigation into analyst needs and is

expected to proceed with detailed system design, and implementation of initial segments over the next three years. The proposed system would cover all document sources of significance to the analyst, all categories of information (persons, organizations, places, things, subjects), and all geographic areas, with minimum handling during processing. It is expected to incorporate input-output devices to speed data transcription and file querying, large capacity memories for index data storage, and a large-capacity document storage system in micro-image form.

Both the Army and the Air Force have sponsored research in automatic abstracting from intelligence documents. Under this concept, the ADP equipment automatically selects and prints significant sentences from a report. Attached as the covering sheet of a report, the Auto Abstract provides a good summary of the full contents.

B. Biographic Intelligence Support

The WALNUT System, now at an advanced stage of development in CIA, is a large file on foreign personalities of counterintelligence interest. Automated document storage and retrieval equipment has been installed, a sophisticated name-searching technique has been designed, and development of very large random access computer memory equipment to store the entire biographic index is at an advanced stage. A proposal is under study to develop a name tracing communications network that would provide quick access to information on foreign personalities stored anywhere in the Intelligence Community.

C. Military Intelligence Support

Targeting, as we know it today, could not exist without computers to store, collate, and retrieve the detailed data on 110,000 targets world-wide. Computers have also proved their worth in assisting in the production of radar order-of-battle from ELINT; in producing air, naval and ground forces order-of-battle; in following current air and merchant shipping activity; in compiling damage assessment information; in intelligence support to war gaming; and in performing various mathematical computations. Efforts are currently underway to extend ADP to such functional areas as Current Intelligence and Indications, Collection Assets Inventories, Ports and Harbors, Coasts and Landing Beaches, Foreign Railroad and Highway Transportation Systems, and Photographic Storage and Retireval.

In the late 1950's the Army began to develop a computer system for collating incoming fragmentary reports on foreign ground forces order-of-battle. Under this system, detailed items of information on foreign military units, e.g., names of unit commanders, unit designations, locations, will be organized and collated to establish full identification and movement of military units. In 1963, this system was transferred to DIA and is currently undergoing testing.

D. Overhead Reconnaissance Support

Computer support to overhead reconnaissance has increased significantly during the past two years in CIA and DIA. Computers are being used in the following areas: (1) determining what to collect; (2) technical support of operations; (3) information and technical support of photographic interpretation. The computer-supported system for information retrieval on targets is an outstanding accomplishment in the information

handling field. It brings pertinent data on each target (what's known and what's needed) to the elbow of the photo interpreter as he examines new photography. New information gained from the readout is promptly incorporated in the record used to establish new target priorities and to support new studies of the given target.

Both the Navy and the Air Force have under development semi-automated tactical photo interpretation systems. These are designed to provide the photo interpreter at a console with the necessary reference material without the attendant bulk that is usually associated with reference materials. The consoles also perform certain standard mathematical computations for the photo interpreter. One of these systems is currently installed and undergoing testing on two of the Navy's aircraft carriers.

E. Machine Translation

All three Military Services and CIA have been sponsoring research in machine translation of foreign language documents. The present level of Government support to machine translation is about $3 million annually. Both general and special purpose equipment is being used for this purpose.

Three current developments are especially significant. First, CIA plans to contract for equipment which should both augment translation capabilities and yield valuable data on an operating machine translation system. Second, the Air Force program has reached the stage where the special equipment involved will be transferred from the contractor's facility to an operational environment at FTD in Dayton, Ohio. This equipment will soon permit machine-assisted translation of 100,000 words of

material per day. Third, a Joint Advisory Group on Automatic Language Processing has been formed, with membership from the National Science Foundation, Department of Defense elements, and CIA. This group has agreed to support jointly a central staff to coordinate Government-supported machine translation research, evaluate its results, and guide its course in order to accelerate progress in this field.

F. Intelligence Support of Command and Control Systems.

A considerable amount of research and development work is currently underway to provide intelligence in machinable form to the various command and control systems. The latter include the entire range of facilities from the National Military Command Center and its Alternates, which will serve National Command Authorities, all the way down to such tactical systems as the Naval Tactical Data System (NTDS) in individual ships. The intelligence provided must be tailored to the decision-making process of the command served over a wide spectrum of emergency conditions.

IV. CRITICAL AREAS IN FUTURE RESEARCH

Three broad areas of research of critical importance to intelligence are being considered: (A) conversion of hard copy material into machinable form; (B) development of massive random access memories; and (C) development of concepts and systems for non-numerical intelligence data structuring and processing. Each of these is discussed briefly below:

A. Conversion of hard copy material into machinable form. At present, the bulk of the intelligence material processed on ADP equipment is first formatted manually from free text, then key punched onto punch cards, verified, converted to magnetic tape,

S-E-C-R-E-T

- 10 -

and finally entered into the computer. This is a costly procedure in terms of money, personnel and time.

For all intelligence ADP applications which entail huge data bases of historical material or which are characterized by voluminous daily inputs, the cost of data conversion is so great as to be a major problem. This is especially true of the processing of data on foreign missile, air, ground, naval, and merchant shipping activity. It is also true of projects pertaining to foreign transportation systems, ports, and harbors, coasts and landing beaches and biographic information. It is apparent that ADP will never be applied to the bulk of the historical or current intelligence material until the data conversion problem is solved. The solution may lie, in part, in research efforts currently underway to develop an optical scanner which can recognize lower and upper case letters and digits, and automatically punch the text on cards or paper tapes. As for the formatting problem, the solution may lie in programming efforts designed to permit the processing in computers of material which is virtually in full text.

(B) Development of Massive Random Access Memories. Because of the limited internal memory capacity of computers, data bases today are stored on punched cards, magnetic tape, disc files or drums. None of these are entirely satisfactory. Storage capacity on cards or magnetic tape is theoretically unlimited, but processing speeds are relatively slow. Processing speeds on disc files or drums are much faster but storage capacity is limited.

S-E-C-R-E-T

Due to the enormous volumes of material encountered in intelligence processing, there is a strong requirement for massive random access memories. The largest typical internal random access computer memory today contains about 300,000 words. Disc memories containing three million words are just becoming commercially available and even larger ones - up to 15 million words - are in the development stage. The DOD is currently sponsoring several projects to develop increased capacity memories.

In order to attain a real breakthrough in this area and achieve, say a multi-billion bit magnetic memory, great strides in batch fabrication techniques for both the magnetic elements and the semi-conductor circuitry are necessary. Superconductive thin film technology offers the possibility of a very large capacity memory because of its ideal storage and switching characteristics combined with its integrated miniature batch fabrication possibilities.

(C) <u>Non-numerical intelligence data structuring and processing.</u>

Three of the most promising non-numerical techniques being pursued at present are adaptive procedures, self-organizing procedures, and content-addressing procedures. All of these are dependent in varying degree on the development of what has come to be known as "associative memory techniques."

Under this concept, the computer selects information on the basis of content, rather than on the basis of location of information or of indexing. All of memory can thereby be interrogated at one time. Such content addressable, or associative, memories could make searching for information very simple and could either simplify

or make unnecessary such tasks as ordering, merging, sorting, and collating of information which require so much time in today's serial processing.

Very little has been written either on associative techniques or on the organization of a data processor which includes an associative memory. At the present time a bibliography of such papers includes less than 20 authors. In turn, very little of what has been written has been directly funded by the Government. Most has been company- or university-sponsored. Current efforts are highly scattered and are not focused on any definable goal. In particular, there are as yet no theories of machine organization or of information processing adequate to guide hardware design efforts. As a result, associative memories are being designed and built with minimum reliance on logical design concepts. Within DOD no projects aimed at development of data processing organization concepts, of techniques, or of applications for associative memories have been funded (although there are several hardware programs for construction of small associative memories). Approximately $1 million is available in FY-64 for research in this area, however.

V. RECOMMENDATIONS

A. Shortage of Research Personnel

Intensification of basic and applied research in ADP techniques of intelligence interest will be dependent primarily on increasing the number of first-rate research workers in universities, laboratories, and private research institutes who are engaged in tasks such as those mentioned in Section IV above. The USIB should express to the National Science Foundation its concern over the acute shortage of skilled research personnel in the field of non-numerical processing and assign CODIB the responsibility

of working with the Foundation to determine what remedial actions might be taken.

B. Interagency Mechanisms for Cooperative Effort

Present-day problems in handling information are not unique to the Intelligence Community. Major investigations are underway in many quarters, particularly in the scientific community. It is likely that vigorous action by the Federal Council on Science and Technology's Committee on Scientific Information could and should contribute in areas of importance to intelligence. The USIB should maintain adequate ties with inter-agency mechanisms engaged in research in information handling techniques applicable to intelligence.

S-E-C-R-E-T